Artist Recognition

Final Project for Machine Learning & Predictive Analytics Class

Linh Le

25th May 2023

Problem Statement

The purpose of this project is to combine the fields of art and science

Project Motivation

Art has been regarded as the pinnacle of intellectual and imaginative expression throughout human history, with painting being a popular medium that connects people across generations. While identifying a favorite artist's work may come naturally to an art enthusiast, it often requires years of practice and research. This project aims to explore whether a machine can achieve the same level of discernment and accurately identify the artist behind a painting. Such identification can help with art appraisals, authentication, provide historical context, and ultimately impacting the artwork's market value.

Project Objective

For this project, we will develop multiple algorithms that identify the artist when provided with a painting and choose the best performing model

Data and Model Hypotheses

For an Image Classification task, we will explore different types of Convolutional Neural Networks

Data Expectations

- **Sufficient data representation:** It is expected that there is enough variation in the data so that the models can capture meaningful patterns and generalizable feature
- **Image quality and consistency:** The images are expected to have consistent quality and resolution
- **Label accuracy:** It is expected that the name of artists assigned to each painting is accurate and reliable
- **Satisfactory class distribution:** The images are expected to be sufficiently evenly distributed across different classes, which mean that artists have appropriately the same number of paintings
- **Input image preprocessing:** Images are to be appropriately preprocessed (i.e. resizing, normalization, augmentation, etc.) as inputs for the models

Model Expectations

- **Sufficient model capacity:** the CNN models need to have appropriate architecture and capacity to capture complex features and patterns from the training data
- **Transferability of learned features:** we will be experimenting with pre-trained CNN models, thus is is expected that the pre-learned features will lead to better performance and faster convergence for the task at hand
- **Appropriate model tuning:** the model's hyperparameters (i.e. learning rate, optimizer) are properly tuned to optimize the training process and enhance the model's ability to generalize

Explanatory Data Analysis – Data Source Description

- Our data comprises of a collection of artworks by the 50 most influential artists of all time in all genres, from Impressionism to Byzantine Art, along with the artists' basic information scraped from Wikipedia.
- **Image Statistics:** There are 8446 painting images, each having different dimensions ranging from 204 pixel to 4096 pixel, totaling 2.4 GB

Painting Image Statistics			
mean	951	**min**	204
sd	141	**max**	4096

- **Class Distribution:** There is an imbalance in number of paintings per artists, with Van Gogh having the most paintings at 877, and Jackson Pollock having the least with 24.

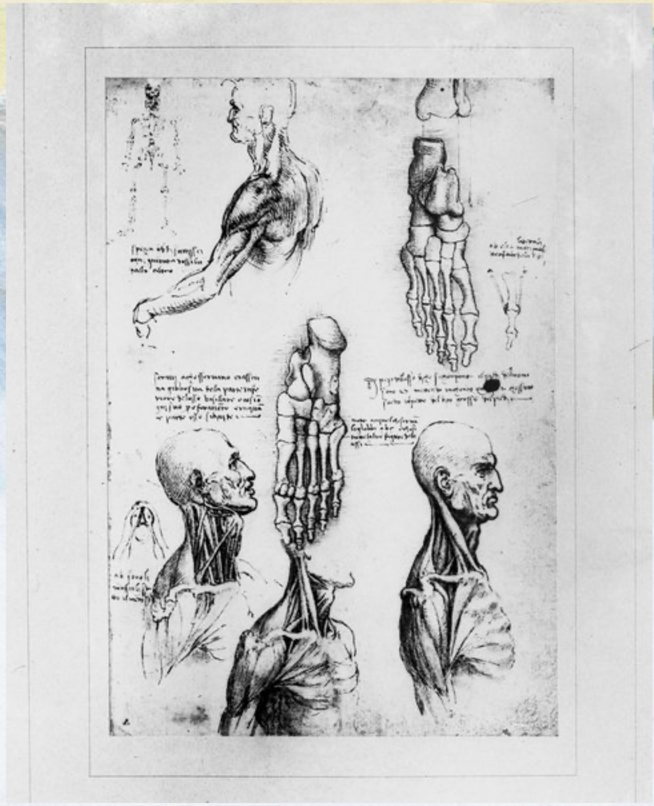
Number of Paintings per Artist Statistics			
mean	169	**max**	877
sd	158	**median**	123
min	24	**75th percentile**	192

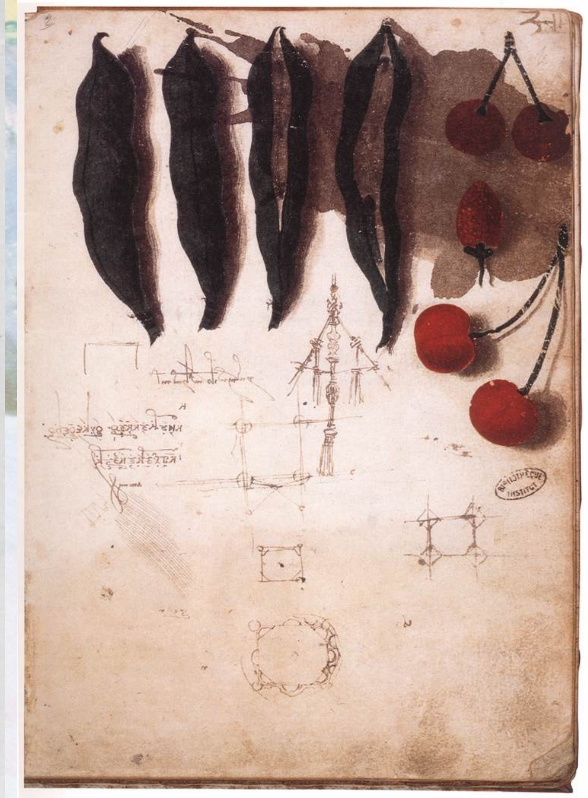
Top 10 Class Distribution

Artist	Number of Paintings
Vincent van Gogh	877
Edgar Degas	702
Pablo Picasso	439
Pierre-Auguste Renoir	336
Albrecht Dürer	328
Paul Gauguin	311
Francisco Goya	291
Rembrandt	262
Alfred Sisley	259
Titian	255

Explanatory Data Analysis – Example Paintings

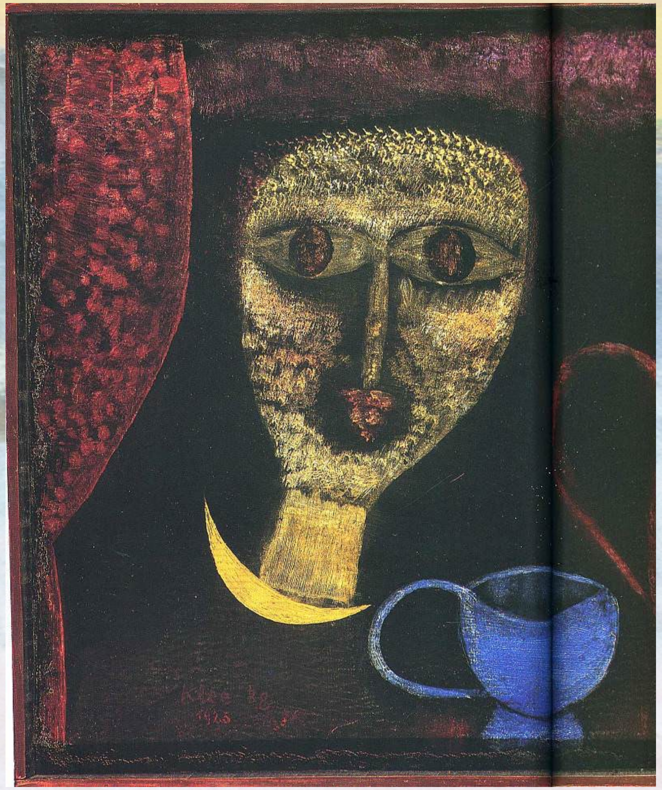
Leonardo da Vinci





Explanatory Data Analysis – Example Paintings

Paul Klee



Explanatory Data Analysis – Example Paintings

Michelangelo

Proposed Approaches

Our project aims to develop a few CNN models for image classification. The proposed approach involves the following steps

- **Data Preprocessing**: The dataset will be preprocessed by resizing images to a fixed dimension, applying data augmentation techniques for increased diversity, and normalizing pixel values between 0 and 1
- **Model Architecture Design**: We will experiment with various architectures, including established ones such as VGG16 and ResNet, and incorporate convolutional and pooling layers, activation functions, and regularization techniques
- **Training**: The model will be trained using the preprocessed dataset. This involves initializing the model, performing forward propagation, defining a loss function, backpropagation for parameter updates, and employing the appropriate optimization algorithm
- **Evaluation**: The trained model will be evaluated using a validation set. The Performance metrics like accuracy, precision, recall, and F1 score will be calculated. Hyperparameter tuning will be performed to optimize performance, and a comparison and analysis of different models will be conducted.

Checks for Overfit and Underfit

To identify underfitting and overfitting in our model, we will have the following checks in place

- **Loss Curves**: We will plot the training and validation loss. High values for both indicate underfitting, suggesting the model fails to capture patterns in the data
- **Accuracy Curves**: We will plot the training and validation accuracy. Low values for both suggest underfitting, indicating the model's inability to learn from the data
- **Model Complexity**: We will assess the model's complexity and capacity. A model that is too simple may lead to underfitting
- **Cross-Validation**: We perform k-fold cross-validation to assess model performance on different data subsets. Consistently poor performance suggests underfitting, while high variability implies overfitting
- **Regularization Techniques**: We will apply techniques such as dropout and early stopping to prevent overfitting and reduce model complexity

Feature Engineering & Transformations

To enable the model to make better predictions, we extract meaningful and discriminative features with data preprocessing and feature engineering

Image Pre-processing

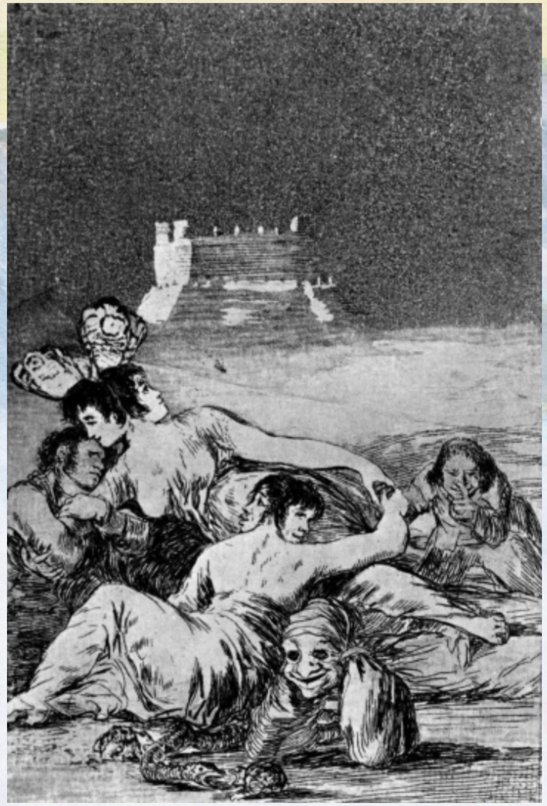
- Image Resizing: All images are resized to 224x224x3 for computational efficiency. This size is chosen based on the consideration that the paintings contain intricate and complex features that can be adequately represented even when scaled down, while still preserving their integrity. Furthermore, 224x224 is a widely adopted size in pre-trained models like VGG and ResNet, which are the models we plan to utilize for our classification task.
- Image Augmentation: *ImageDataGenerator* was used to augment data to provide the model with additional variations of the input images, which can help improve generalization. It:
  - Normalizes images: by setting rescale = 1./255., the pixel values are scaled to the range [0,1]
  - Shears images: the shear angle will be randomly chosen between -5 and 5 degrees during data augmentation
  - Flips images horizontally randomly
  - Flips images vertically randomly

Feature Engineering

- Filter out artists having less than 123 paintings: in order to mitigate the problem of gross class imbalance, all artists with number of paintings less than the median (123) were dropped from this analysis. This will also help with the problem of limited computational resource we are facing
- *Class weight* feature: The class weight is calculated as the sum of all the paintings divided by the product of the number of artists and the number of paintings for each artist. This calculation is intended to assign higher weights to artists with fewer paintings, aiming to balance the impact of different class frequencies during model training

Examples of Original and Transformed Painting

An original & transformed image of Francisco Goya



An original & transformed image of Peter Paul Rubens

Baseline Model Architecture and Results

We first build a simple CNN model as our baseline model, then use various established architectures such as VGG16 and ResNet

Model Architecture

- As our baseline model, a simple CNN was built with 8 layers: Sequential, MaxPooling2D, ConV2D, MaxPooling2D, Flatten, Dense with *relu* activation function, Dense with *softmax* activation function
- The model was trained with a categorical cross entropy loss function, with accuracy as evaluating metric
- The parameters learning rate and optimizer were hyper-tuned to find the best combination for this dataset. The best learning rate of 0.0001 and best optimizer Adam were then used for the followed model architectures

***Table**: Baseline Model Architecture*

```
Model: "sequential"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 conv2d (Conv2D)             (None, 222, 222, 32)      896

 max_pooling2d (MaxPooling2D  (None, 111, 111, 32)     0
 )

 conv2d_1 (Conv2D)           (None, 109, 109, 64)      18496

 max_pooling2d_1 (MaxPooling  (None, 54, 54, 64)       0
 2D)

 flatten (Flatten)           (None, 186624)            0

 dense (Dense)               (None, 128)               23888000

 dense_1 (Dense)             (None, 11)                1419

=================================================================
Total params: 23,908,811
Trainable params: 23,908,811
Non-trainable params: 0
```

Results

The result is promising but still has room for improvement, which we will further explore in the next models

***Table**: Baseline Model Results*

Model Name	Test Loss	Test Accuracy
Baseline CNN	3.72	53.57%

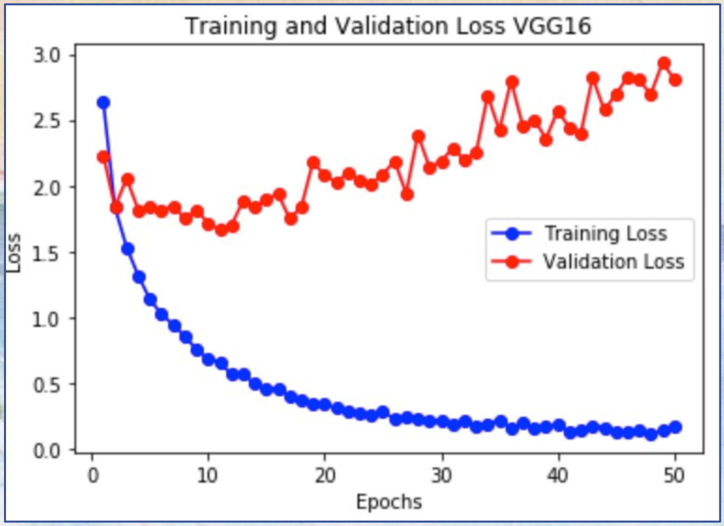
VGG16 Model and Results

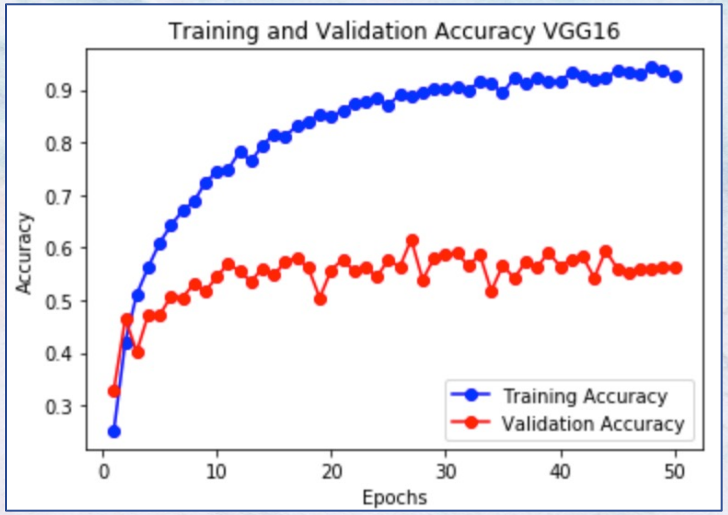
The VGG16 model was then chosen due to its deep architecture, which allows it to capture intricate patterns such as those in our images

- The VGG16 model is commonly used for image classification due to its deep architecture, strong performance on benchmark datasets, and its suitability for transfer learning
- **Model Architecture:** a VGG16 model was built with 5 layers: Sequential, VGG16, Flatten, Dense with *relu* activation, the last Dense layer with *softmax* activation and 25 units as 25 classes to predict
- **Result:** On our dataset, VGG16 performs slightly better than our baseline model, but there were signs of overfitting such as the training loss continues to decrease while the validation loss increases, or the training accuracy continues to increase while the validation loss plateaus

***Table**: VGG16 Results*

Model Name	Test Loss	Test Accuracy
VGG16	2.92	55.24%





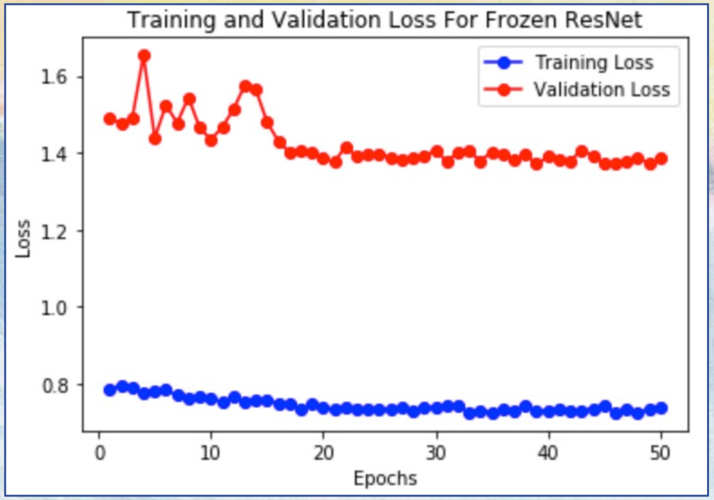
ResNet Model with Frozen Layers Architecture and Results

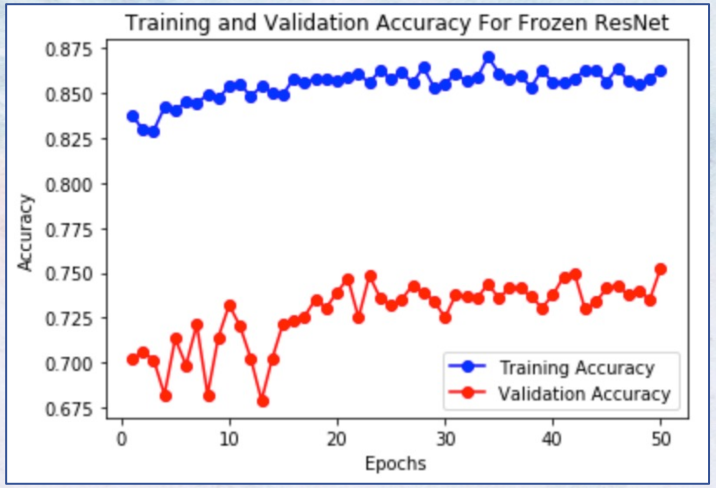
To mitigate the overfit problem that VGG16 had and to improve on the model's ability to capture more complex features and enables better performance, we will explore ResNet with regularization techniques on our dataset

- We choose to employ the ResNet model because it is helpful for tasks involving large-scale datasets and complex image recognition, such as the vast number of high-resolution images in our dataset
- **Model Architecture:** We imported ResNet50 with *imagenet* weights and without the top layer. The model architecture includes the ResNet base model, a Flatten layer, a Dense layer, a Dropout method of 0.5, a Batch Normalization layer with a *relu* activation, another Dense layer, another Batch Normalization with a *relu* activation. The last layer is a Dense layer with 25 units with a *softmax* activation function. For the first run, ResNet layers are frozen so that their weights are not re-trained.
- **Result:** This model yields a breakthrough in our result, having an increase of 17 percentage point in accuracy, and ~51% improvement in test loss than VGG16's

***Table**: Frozen ResNet Results*

Model Name	Test Loss	Test Accuracy
Frozen ResNet	1.40	72.85%





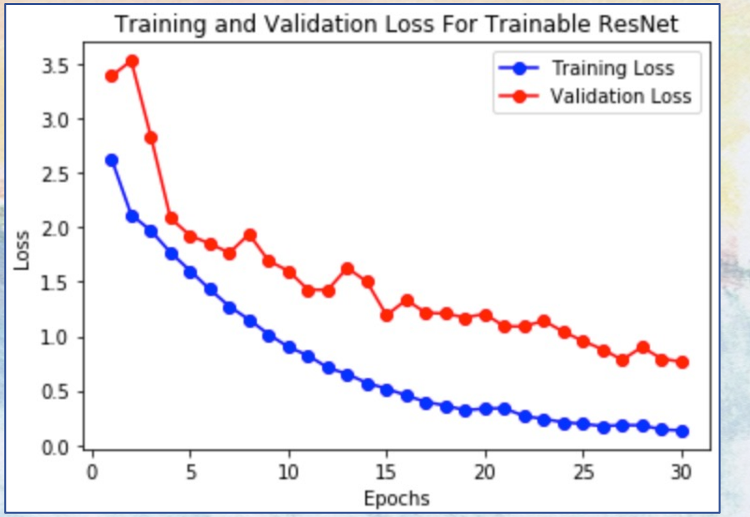
ResNet Model with Trainable Layers Architecture and Results

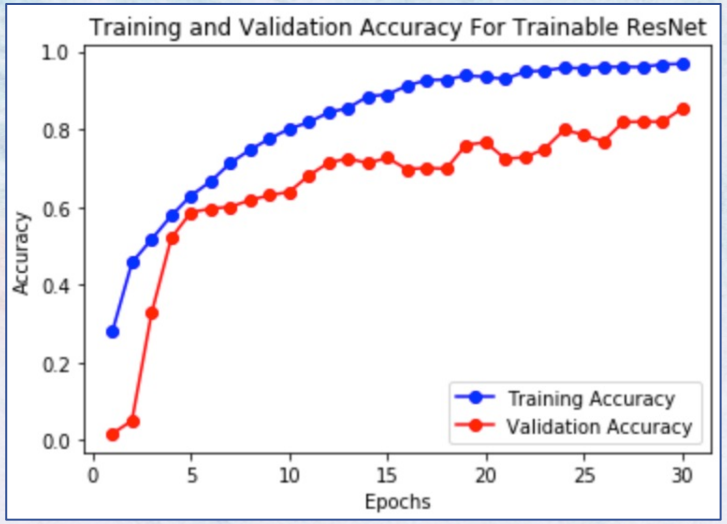
We then explored using ResNet with trainable layers for the task at hand

- In order to evaluate the impact of retraining ResNet layers on the model's prediction accuracy, we maintained the same model architecture as before but modified the condition to allow ResNet layers to be trainable
- However, due to computational limitations, the kernel kept crashing at 50 epochs. To address this issue, two regularization techniques, namely early stopping and reduced learning rate, were implemented. Despite the inclusion of these regularization methods, the kernel continued to encounter memory constraints, preventing completion of a full training session
- In order to ensure the successful completion of training, we decided to reduce the number of epochs to 30. This adjustment allowed the model to train without encountering any memory constraints or interruptions.
- **Result:** This model has the best performance out of all models we have tried, having almost a 50% improvement in test loss and 9 percentage points more than Frozen ResNet

***Table**: Trainable ResNet Results*

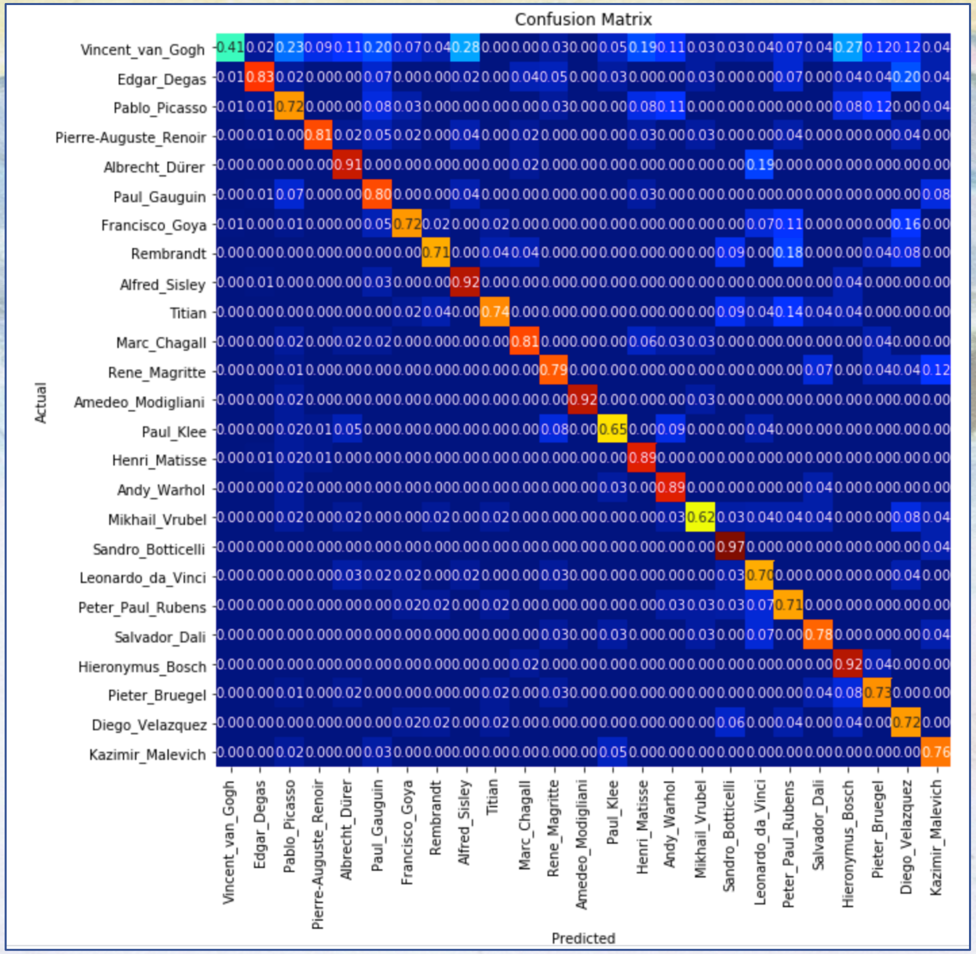
Model Name	Test Loss	Test Accuracy
Trainable ResNet	0.76	81.67%





Correlations and Confusion Matrix of Trainable ResNet Model

The artists that have high correlations are those that the model thinks have the most similar painting styles, thus more prone to making prediction mistakes. For example: Van Gogh and Alfred Sisley, or Edgar Degas and Diego Velazquez.



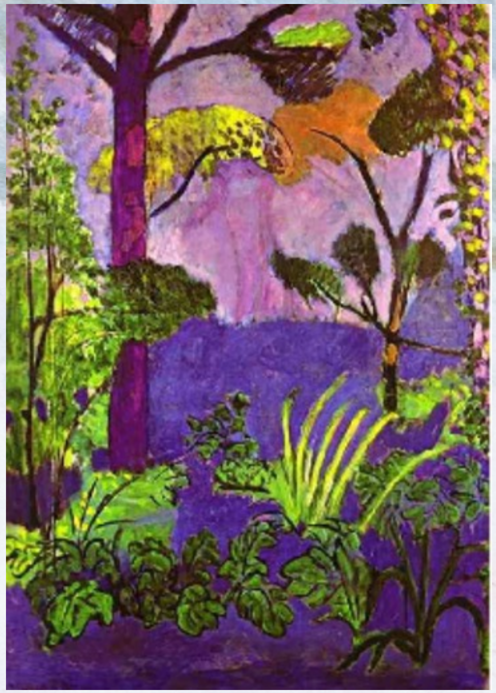
Classification Report:

	precision	recall	f1-score	support
Vincent_van_Gogh	0.96	0.41	0.57	172
Edgar_Degas	0.93	0.83	0.88	140
Pablo_Picasso	0.58	0.72	0.65	87
Pierre-Auguste_Renoir	0.87	0.81	0.84	67
Albrecht_Dürer	0.78	0.91	0.84	64
Paul_Gauguin	0.60	0.80	0.69	61
Francisco_Goya	0.79	0.72	0.76	58
Rembrandt	0.82	0.71	0.76	51
Alfred_Sisley	0.70	0.92	0.79	50
Titian	0.84	0.74	0.79	50
Marc_Chagall	0.84	0.81	0.83	47
Rene_Magritte	0.75	0.79	0.77	38
Amedeo_Modigliani	1.00	0.92	0.96	37
Paul_Klee	0.77	0.65	0.71	37
Henri_Matisse	0.70	0.89	0.78	36
Andy_Warhol	0.69	0.89	0.78	35
Mikhail_Vrubel	0.75	0.62	0.68	34
Sandro_Botticelli	0.72	0.97	0.83	32
Leonardo_da_Vinci	0.56	0.70	0.62	27
Peter_Paul_Rubens	0.51	0.71	0.60	28
Salvador_Dali	0.75	0.78	0.76	27
Hieronymus_Bosch	0.62	0.92	0.74	26
Pieter_Bruegel	0.63	0.73	0.68	26
Diego_Velazquez	0.49	0.72	0.58	25
Kazimir_Malevich	0.63	0.76	0.69	25
accuracy			0.74	1280
macro avg	0.73	0.78	0.74	1280
weighted avg	0.78	0.74	0.74	1280

Predictions of Photos in the Training Set

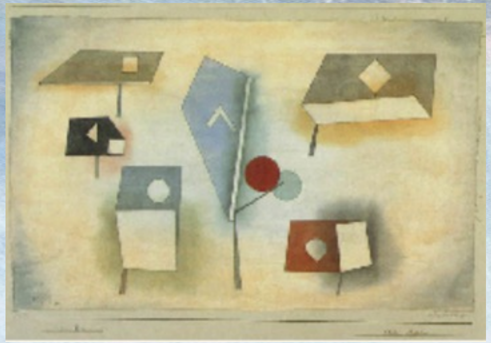
The model demonstrates excellent performance in accurately identifying the artists associated with images in the training set

Actual Artist	Predicted Artist	Prediction Probability
Henri Matisse	Henri Matisse	75.86%



Actual Artist	Predicted Artist	Prediction Probability
Albrecht Dürer	Albrecht Dürer	52.84%

Actual Artist	Predicted Artist	Prediction Probability
Titian	Titian	42.83%



Actual Artist	Predicted Artist	Prediction Probability
Paul Klee	Paul Klee	30.68%

Predictions of Photos not in the Training Set

The model exhibits a slightly lower performance, with lower confidence when dealing with images that are not present in the original dataset

Actual Artist	Predicted Artist	Prediction Probability
Paul Klee	Paul Klee	44.36%

***Image URL**: https://uploads0.wikiart.org/images/paul-klee/flower-myth-1918(1).jpg!Large.jpg*

Actual Artist	Predicted Artist	Prediction Probability
Rene Magritte	Rene Magritte	43.56%

***Image URL**: https://www.menil.org/collection/objects/4901-golconda-golconde*

Actual Artist	Predicted Artist	Prediction Probability
Pablo Picasso	Henri Matisse	26.92%

***Image URL:** https://art.branipick.com/portrait-of-angel-fernandez-de-soto-picasso-1903-1509-x-1950/*

Learnings from the Methodology

Based on the comparison of trainable and untrainable versions of CNN, VGG16, and ResNet models, several key learnings were derived

1. **Transfer Learning Effectiveness:** The performance of transfer learning, where pre-trained models are utilized as a starting point, varies across different architectures. It is observed that ResNet, being a deeper and more complex model, tends to benefit more from transfer learning compared to CNN and VGG16
2. **Importance of Model Architecture:** The choice of model architecture significantly impacts the task's performance. ResNet, with its skip connections and residual blocks, demonstrates superior performance due to its ability to mitigate the vanishing gradient problem and capture more intricate features
3. **Retraining vs. Frozen Layers:** Fine-tuning or retraining specific layers in pre-trained models can be advantageous in certain cases. For instance, allowing ResNet layers to be trainable leads to improved prediction accuracy, indicating that some task-specific features can be learned by adapting the pre-trained model
4. **Computational Constraints:** The computational limitations should be taken into consideration when designing the methodology. It is essential to find a balance between model complexity and available resources to ensure successful training and avoid crashes or memory issues
5. **Regularization Techniques:** The inclusion of regularization techniques, such as early stopping and reduced learning rate, can help mitigate overfitting and improve model generalization. However, in cases where memory constraints persist despite regularization, alternative approaches or compromises may be necessary

Future Work

Some possible future work that addresses both business needs and model improvement to enhance the practicality, applicability, and overall performance of the model in real-world business scenarios

1. **Domain-Specific Feature Incorporation**: In order to optimize performance for business applications, integrating domain-specific features into the model is essential. Our future plans involve expanding this project into an art appraisal tool, leveraging the existing model while incorporating additional information such as artist details, historical prices, and painting year. This enhancement aims to improve the model's relevance and predictive accuracy within the art appraisal domain
2. **Model Robustness Evaluation:** We can further assess the model's robustness by subjecting it to diverse and challenging datasets, such as OOD dataset, noisy dataset, or adversarial examples
3. **Real-World Testing:** The model's performance can be validated for its scalability, and reliability through real-world testing and evaluation in practical business settings
4. **Collaboration with Domain Experts:** Collaboration with domain experts and stakeholders can help us gain valuable insights, refine the model, and better align it with specific business requirements
5. **Quantify Business Impact**: The actual business impact should always be measured and quantified, assessing its effectiveness in terms of cost savings, revenue generation, and other relevant metrics


Thank You